DETECTION OF MOVING OBJECT BY FUSION OF COLOR AND DEPTH INFORMATION

T. T. Zhang， G. P. Zhao and L. J. Liu

School of Automation

Nanjing University of Science and Technology, Private bag 210014

Nanjing, China

Emails: zhaogaopeng@sina.com

Submitted: Dec. 1, 2015 ***Accepted: Jan. 24, 2016*** ***Published: Mar. 1, 2016***

Abstract- Moving object detection based on color information is easily affected by illumination changes and shadows in complex scenes. Depth information can provide complementary information. In the paper, a novel method is presented by using color and depth information. Firstly, we improve the codebook algorithm by fusing the depth information as the fourth channel in the code word. Next, a compensation factor algorithm is presented to make the edges accurate. So the final detection result can be obtained by logic operation. Experiments adapt the public datasets, and experimental results show that the proposed method can successfully cope with the limitations of color-based or depth-based detection.

Index terms*:* Object detection, codebook, edges, color information, depth information.

I. INTRODUCTION

With the development of information technology, there have been widely applications of computer vision such as video surveillance, human computer interaction and video conference[1-2].Extracting moving foreground objects from a video sequence is the first critical step in video analytics[3-4]. The most widely used approach of moving object detection is background subtraction. The fundamental of background subtraction is that the moving objects detection obtained from the difference operation between the current frame and reference background model, which is the existing background model or re-established a background model through statistical modeling. There are many algorithms based on color information in the literature about background subtraction, such as Mixture of Gaussians (MOG) [5], non-parametric kernel density estimation[6] and Codebook (CB) [7]. However, on the one hand, above color-based algorithms face many challenging problems including the following: vulnerable to illumination changes; shadows cast by moving objects; camouflage(i.e., similar color between moving objects and the background)[8]. On the other hand, MOG faces a problem that the learning rate which is one of the important parameters to adapt to background changes is difficult to adjust. If the model adapts too slowly, sudden change to the background cannot be detected in a wide model. For a high learning rate, background model will absorb foreground pixels which are moving slowly[9]. Although non-parametric kernel density can quickly adapt to changes in the background process, it has high requirements for hardware memory due to large amount of calculation[10].

In recent years, many authors have come up with some methods that combine color and depth information to detect moving object. In [11], a logical “or” is used to combine the different foregrounds that respectively come from grayscale image and distance image. Although the method cope with problems that there are similar color and closed distance between the objects and background, but it fails to overcome the edge noise. In [12], the author developed the background subtraction based on the Gaussian Mixture Models using color and depth information. It not only solved the limitation of color camouflage, but also decreased the depth noise. However, it is difficult to deal with complex scenes.

In this paper, we present a novel background subtraction algorithm which is referred as $CB_{CD\&E}$. Firstly, depth information is not only as the fourth channel on the codebook algorithm, but also is as the further condition when judging a pixel belongs to the foreground or background, which is referred as $CB_{C\&D}$. Secondly, a compensation factor algorithm is introduced to make the edge of detected objects accurate, and a series of logical operation are used to generate the final result. The results show a quantitative and qualitative improvement in the moving object detection.

II. PROPOSED METHOD

The overview of the proposed method is presented inFigure1. There are two stages of the proposed algorithm.

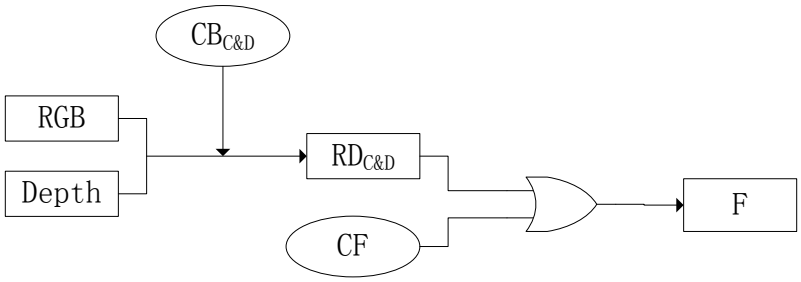


Figure 1. Schematic of object detection algorithm: CF stands for compensation factor. F stands for final foreground mask.

In the stage 1, depth information, which is as the fourth channel (R, G, B, D) of codebook algorithm, enhances the conditions of background construction and foreground detection. In stage 2, the final output F is obtained by a logical "or" operator between $RD_{C\&D}$ obtained in stage 1 and compensation factor CF. CF that is acquired through a series of logical operations strengthen the edge detection. The logical "or" operation between $RD_{C\&D}$ and CF is defined as follows:

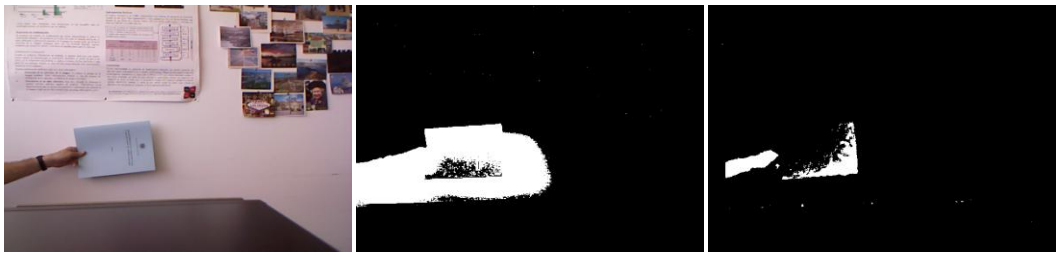
$$F^i = RD^i{}_{C\&D} \vee CF^i \quad (1)$$

Among them, $i = 1...n$ with n total number of frames. The following will detail the above stages.

a. Codebookbased on color and depth ($CB_{C\&D}$)

Depth-based algorithm has strong robustness on sudden lighting changes, highlighted regions and shadows, which color-based algorithm cannot overcome. Actually, we can only use the depth-based codebook algorithm to remove shadows and highlighted. However, when the objects are

closed to the background, the pixels will be classified as the background. In figure2, we can see that there are shadows in color-based algorithm; moreover, foreground objects are mistakenly detected based on depth algorithm due to the close range between the objects and background. Consequently, not only do we use the depth as the fourth channel, but also we fuse the color and depth information to detect foreground.



(a) test frame (b) CB_{Color} (c) CB_{Depth}

Figure 2. The example of complicated scenario;(a) is the test frame; (b) is the result of color-based codebook algorithm; (c) the result of depth-based codebook algorithm

The conditions contain not only the color and brightness distortions, but also the depth disparity. The approach, which determines whether one-dimensional depth pixel matches the codeword, is in a similar way to the brightness. $D_{\min}$ and $D_{\max}$ respectively represent the minimum and maximum depth values in a codeword. Depth distortion is allowed to vary in a certain range, [$D_{low}, D_{\max}$], defined as:

$$\begin{cases} D_{low} = \alpha_D D_{\max} & 0.4 \le \alpha_D \le 0.7 \\ D_{hi} = \min\left\{ \beta_D D_{\max}, \dfrac{D_{\min}}{\alpha_D} \right\} & 1.1 \le \beta_D \le 1.5 \end{cases} \quad (2)$$

Therefore, each codeword c_i includes not only $v_i = (\overline{R_i}, \overline{G_i}, \overline{B_i}, \overline{D_i})$ but also an eight-tuple $aux_i = \langle I^i_{\min}, I^i_{\max}, D^i_{\min}, D^i_{\max}, f_i, \lambda_i, p_i, q_i \rangle$. The logical disparity function is defined as follows:

$$disparity(D, \langle D_{\min}, D_{\max} \rangle) = \begin{cases} true & if \ \neg Valid(D) \vee (D_{low} \le D \le D_{hi}) \\ false & otherwise \end{cases} \quad (3)$$

Compared with color distortion of original Codebook, we add depth information as further conditions in $CB_{C\&D}$. If the color distortion is less than the threshold ε_1 or color distortion is between ε_1 and ε_2 ($\varepsilon_2 = 1.6\varepsilon_1$),meanwhile, $disparity(D, \langle D_{\min}, D_{\max} \rangle)$ is true, The $color(x)$ in $CB_{C\&D}$ is true. $color(x)$ is defined as follows:

$$color(x) = \begin{cases} true & if\ colordist(x, c_m) \le \varepsilon_1 \vee \\ & \vee (\varepsilon_1 < colordist(x, c_m) \le \varepsilon_2 \wedge disparity(D, \langle D_{\min}, D_{\max} \rangle)) \\ false & otherwise \end{cases} \quad (4)$$

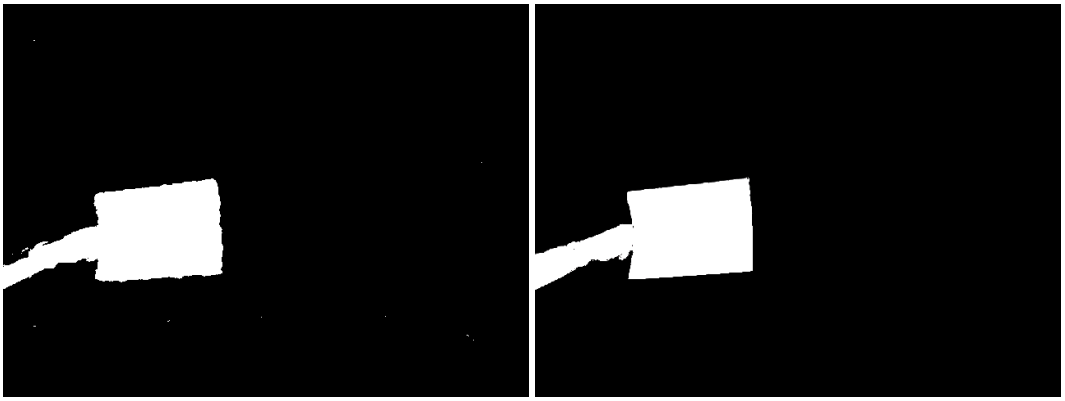
The significance of the formula 5 is to determine those critical pixels belong to the foreground or the background. Due to the input pixel is close to the threshold, using based-color algorithm cannot accurately detect it. Therefore, the depth value of that pixel is considered as further condition.

In the end, the algorithm matches the current pixel value with the appropriate codeword according to the computed color, brightness and disparity distortions. $BGS(x)$ is defined as follows:

$$BGS(x) = \begin{cases} BG\ \ if\ color(x) \wedge brightness(I, \langle I_{\min}, I_{\max} \rangle) \\ \quad \wedge disparity(D, \langle D_{\min}, D_{\max} \rangle) \\ FG\ \ otherwise \end{cases} \quad (5)$$

b. Compensation factor and fusion

In the stage 1, we have evaluated $CB_{C\&D}$ that reduces the impact of that closed distance between object and background in the resultant segmentation without having to perform uneven edges. There is a comparison between the test frame of $CB_{C\&D}$ and the ground truth obtained by manual segmentation in Figure 3. It can be observed in Figure 3(a) that the edges are not accurate.



(a) $CB_{C\&D}$ (b) the ground truth

Figure 3. The comparison of the test frame in Figure 2; (a) is the result of the $CB_{C\&D}$ method; (b) is the ground truth

In order to make the edges more accurate, Compensation factor algorithm CF is designed and its framework is shown in Figure 4.

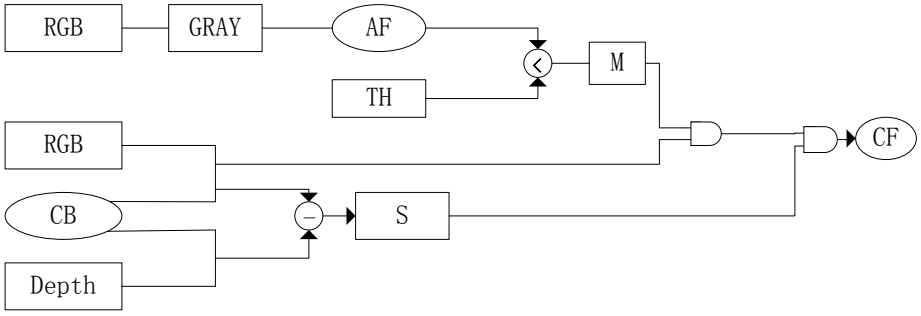


Figure 4.Schematic of compensation factor based on depth information algorithm: S stands for the subtraction mask. AF stands for averaging filter. TH stands for the decision threshold in pixel unit. M stands for depth enhanced mask.

There are four main steps in the compensation factor algorithm.

(1)Averaging filter

First of all, we convert RGB images to gray images. Then, the average gray levels of each pixel, which is in a neighborhood that pixels are surrounding in a square window, take the place of original gray levels of each pixel by using averaging filter. $T^i(j,k)$ ($i=1...n$ with n total number of frames) represent the filtered images that a generic pixel coordinates in (j,k), and M stands for the half-side of the squared kernel.

$$T^i(j,k)=\frac{1}{4M^2}\sum_{s=-M}^{M}\sum_{t=-M}^{M}GRAY^i(s,t) \quad (6)$$

(2)Threshold determination

In this step, our goal is to obtain a logical depth-enhanced mask M^i by comparing the intensity value between $T^i(j,k)$ and $GRAY^i(j,k)$. $M^i(j,k)$ is defined as follows:

$$M^i(j,k)=\begin{cases}1 & for\ \left|GRAY^i(j,k)-T^i(j,k)\right|<TH\\ 0 & otherwise\end{cases} \quad (7)$$

A logical depth-enhanced mask M^i, which is a criterion of choosing pixel in the subtraction mask S^i to get CF, can be completed by repeating iteratively above-mentioned step for each pixel.

(3)Obtain subtraction mask S^i

To obtain subtraction mask S^i, we perform a subtraction operation between color FG frames and depth FG frames, which are binary images.

$$S^i = Depth^i_{CB} - RGB^i_{CB} \quad (8)$$

Our purpose, which is to obtain missing edge for the original mask RD_{DECB}, leads us to consider only positive pixel value in S^i.

$$S = \begin{cases} S^i & if\ S^i(j,k) > 0 \\ 0 & if\ S^i(j,k) < 0 \end{cases} \quad \forall (j,k)\ j = 1...J; k = 1...K \quad (9)$$

In Equation (9), (j,k) are the coordinates of a generic pixel, J and K stand for the total number of rows and columns respectively. The approach is specially aimed to make a logical mask contain regions of $Depth^i_{CB}$ without RGB^i_{CB}.

⑷Two logical "and" operation

If a pixel comes from a uniform color region, we consider it valid. In that way, its level of intensity is really similar to the medium level of intensity computed in the surrounding averaging window. CF is defined as follows:

$$CF^i = M^i \wedge S^i \quad (10)$$

III. EXPERIMENTAL RESULTS AND ANALYSIS

Five different approaches have been studied and evaluated with the publicly dataset. These approaches are the following ones: the original color-based Codebook(CB_{Color}), the Codebook based only on depth(CB_{Depth}), the 4D Codebook(CB4D) which the depth information is only as the fourth channel in codeword without bias over color threshold, the Codebook based on color and depth($CB_{C\&D}$), and the proposed method($CB_{CD\&E}$). Experiments based on dataset compare the proposed algorithm with the other four methods in three different scenes, such as the situation that object gradually keeps away from the background, the situation that color of object is the same as the background and the situation of sudden illumination changes. On the one hand, the use of a dataset with ground truth segmentation is required to perform a quantitative analysis in addition to the qualitative one; on the other hand, we conducted a quantitative analysis of PR value. Simulation of experiments is finished in VS2010 environment, while taking advantage of the OpenCV library to assist image processing. The test dataset comes from the website(http://atcproyectors.uqr.es/mvision/).

a. Parameter Settings

There is a plurality of parameters that affect the effectiveness of the algorithm in together. In order to achieve good overall performance, we have chosen an optimal set of parameters based on a large number of experiments. According to trial and error, we determine the threshold TH (in 2.2 section) is an empirical value. Table 4.1 shows the values of these parameters:

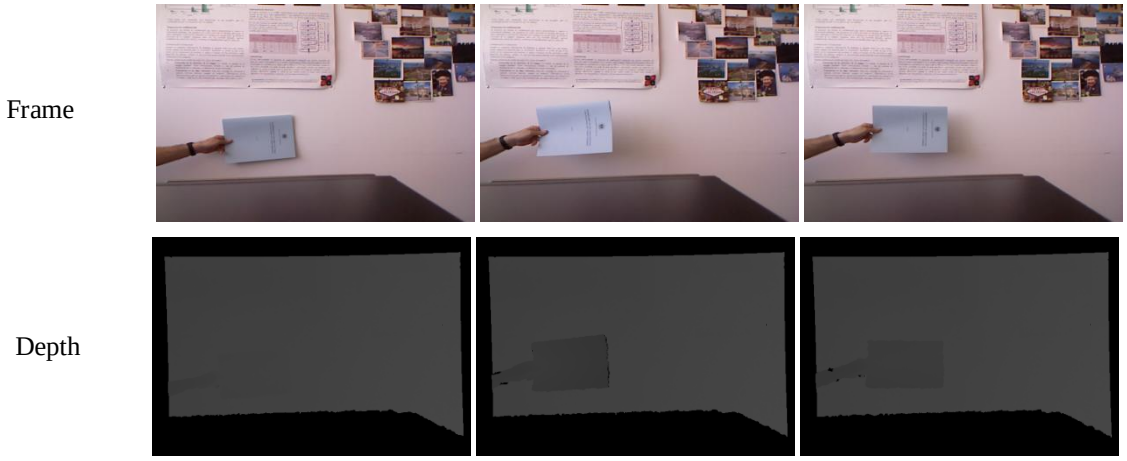
Table 1: Parameters selected for the proposed method

Parameter	Value	Parameter	Value
ε_1	10	TH	20
α	0.75	β	1.3
α_D	0.65	β_D	1.25

b. Qualitative Analysis

b.i Target and Background Similar Distance

Figure 5 shows the qualitative results in the Wall sequence of test dataset, which a person hands a book from near and far, from far and near. We select the 74th frame, the 93th frame and the 134th frame for test respectively, among them the 74th frame is nearest from the background and the 93th is furthest from the background.



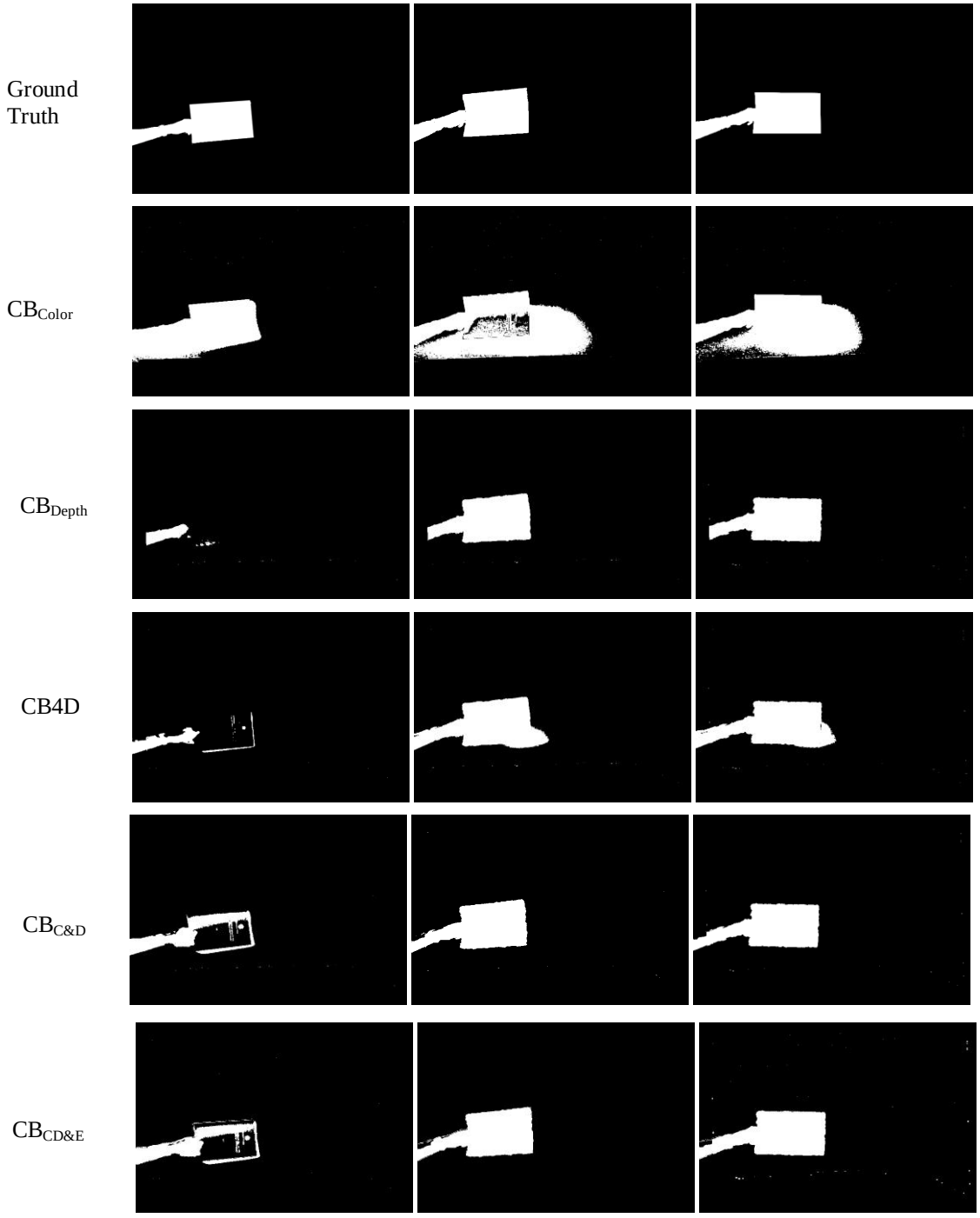
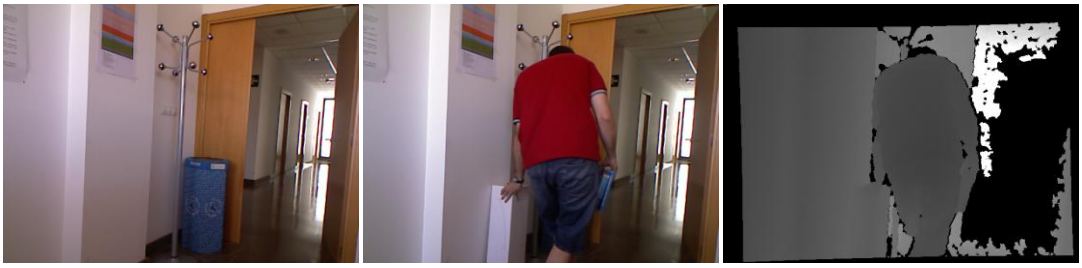


Figure.5 Results obtained from the wall sequence

As can be seen from the above, there are large area shadow in the CB_{Color} because of the similar distance between object and background. CB_{Depth} can solve this problem, but object cannot be detected in the 74^{th}frame. $CB_{C\&D}$ shows good effect when fuse the color information and depth information. The $CB_{CD\&E}$ reduces the noise of edge and makes edge smoother. So $CB_{CD\&E}$ gets the best result.

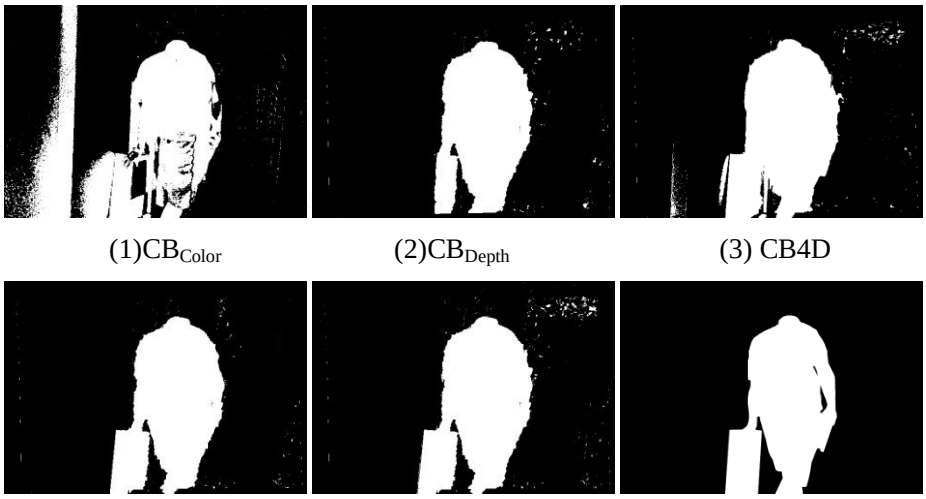
b.ii Target and Background Similar Color

In this section, we choose the Hallway sequence for qualitative analysis in the test dataset. As shown in Figure 6, a man holding white box goes through the background scenes. The color of box is similar to the wall. In order to preferably analyze the results, we only capture the white box portion of the figure in this paper. In Figure 7, we can find the object contained a black shadow cannot be completely detected in CB_{Color}. Although the detection result is better in CB_{Depth}, white box is only partially detected. To compare the CB_{Depth} and $CB_{C\&D}$ in Fig.7, the effect of $CB_{C\&D}$ is obvious. $CB_{CD\&E}$ makes the edge more clearly.



(a) Background (b) 258thFrame (c) 258thDepth

Figure6. The original image



(1)CB_{Color} (2)CB_{Depth} (3) CB4D

(4)$CB_{C\&D}$ (5)$CB_{CD\&E}$ (6)Ground Truth

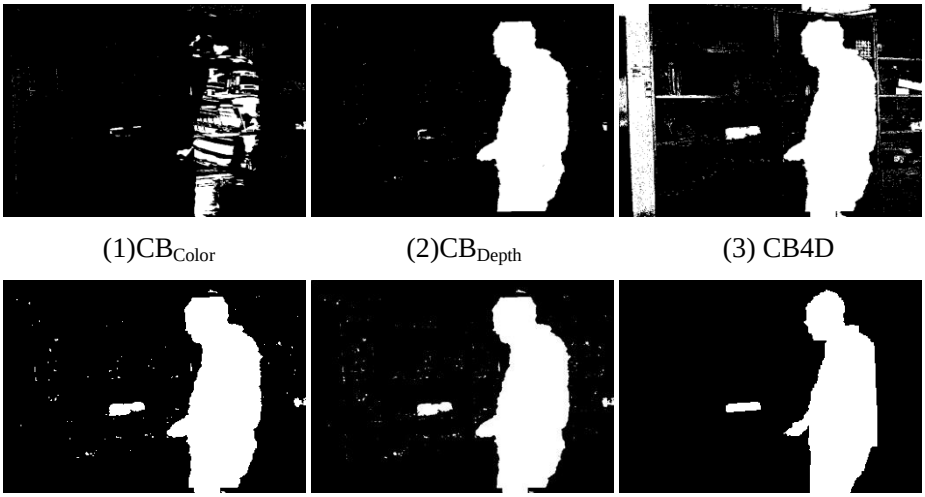
Figure 7. The result of 258th

b.iii Sudden Illumination Changes

Figure 8 and Figure 9 show the qualitative analysis in the Shelves sequence of test dataset. As shown in Figure 8, there are changes of exposure when 365th frame switch to 366th frame. In Figure 9, we can find that CB_{Color} cannot adapt to sudden changes in illumination. There are a lot of noises in CB4D. Most of this noise is filtered by the $CB_{C\&D}$ by means of the fusion method.

(a) 365th Frame (b) 366th Frame

Figure 8. The original image



(1)CB_{Color} (2)CB_{Depth} (3) CB4D

(4)$CB_{C\&D}$ (5)$CB_{CD\&E}$ (6)Ground Truth

Figure 9. The result of 366th

c. Quantitative analysis

precision (P), *recall* (R) and F are the common evaluation criteria of object detection. *recall* is the true positive; *precision* is the ratio between the number of correctly detected pixels and the total number pixels marked as foreground; F -number is a successful combination

of P and R, to comprehensively evaluate the performance of the algorithm. They are defined as follows:

$$P = \frac{TP}{TP + FP} \qquad (11)$$

$$R = \frac{TP}{TP + FN} \qquad (12)$$

$$F = 2 \times \frac{P \times R}{P + R} \qquad (13)$$

TP (True Positive) is the number of pixels as the moving object is correctly detected. TN (True Negative) is the number of pixels to be detected as the background. FN (False Positive) is misidentified as the background pixels. FN (False Negative) is the number of pixels to be mistaken for the moving object. This measure not only offers a trade-off between the ability of an algorithm to detect foreground and background pixels, but also provides a general evaluation of robustness of the algorithm. In general, the value of this estimator is higher, the better the performance. In Figure 10 to Figure 12, curve shows the F-number of evaluation frames in different scenes. As we can see, the algorithm performance which is only based on color information or depth information to detect moving object decline or even fail. The method fusion color and depth information gets better effect in different complex environment. Of course, late fusion makes the algorithm performance more stable.

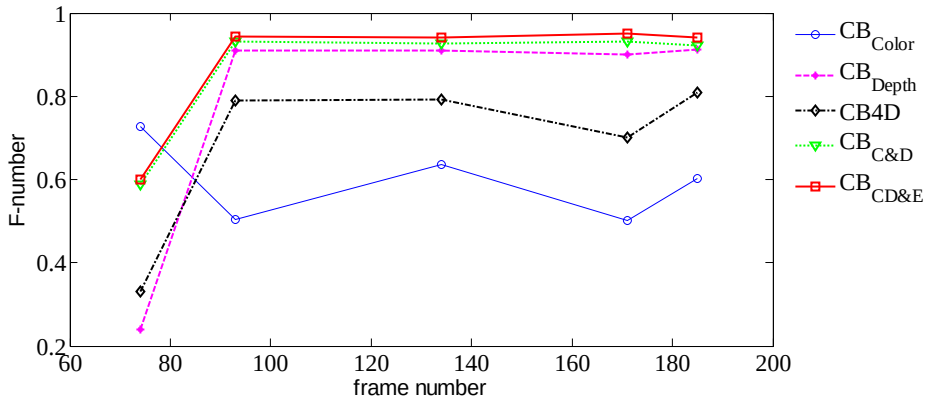


Figure 10.The scene of similar distance

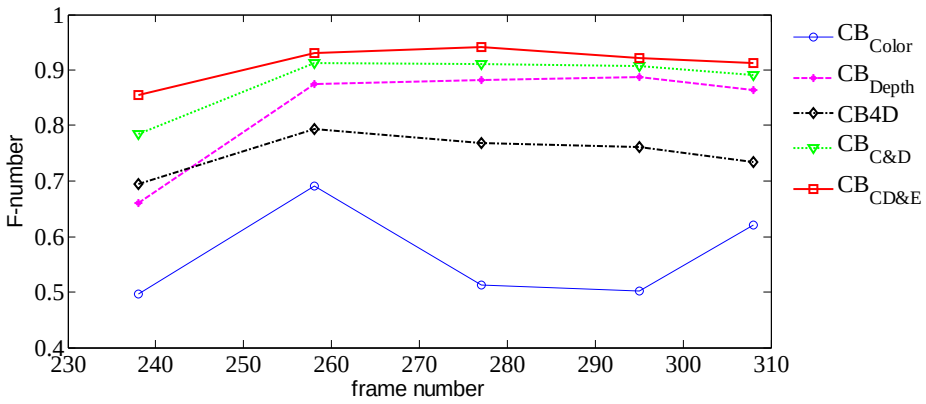


Figure 11.The scene of similar color

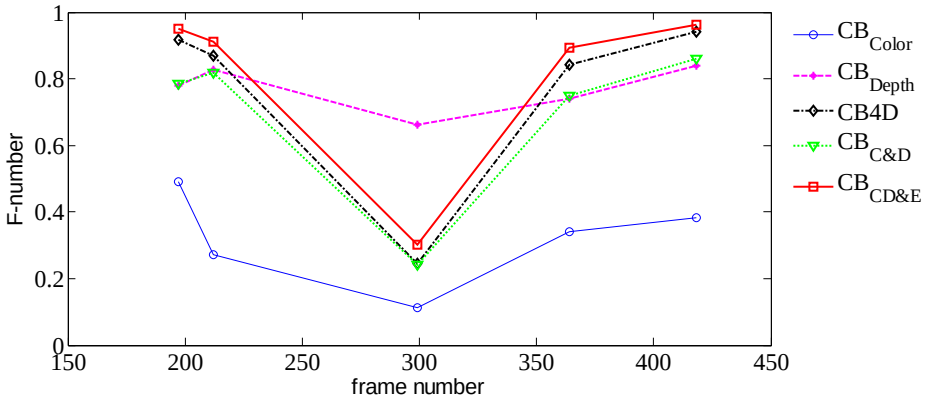


Figure 12. The scene of sudden illumination changes

IV. CONCLUSIONS

The method based on color information cannot cope with classic issues such as the similar color of object and background, sudden illumination changes, and shadow interference. In addition, there will be a leak due to the closed distance in the method based on depth information. In this work, we focus on the combined use of depth and color to reduce the impact of above problems. On the one hand, depth information is as the fourth channel of the codebook and as the further condition of foreground detection; on the other hand, the compensation factor further strengthen the link between color and depth information so that the edge is sharper. Multiple sets of experimental results show the proposed method has higher detection performance in a variety of complex environments.

ACKNOWLEDGEMENTS

This Project is supported by the National Natural ScienceFoundation of China (Grant No. 61203266).

REFERENCES

[1] X.J.Wang, F.Pan and W.H.Wang, “Tracking of moving target based on video motion nuclear algorithm”, International Journal on Smart Sensing and Intelligent Systems,vol. 8, No. 1, 2015, pp. 181-198.

[2] Y.Q.Wang, Y.Z.Zhang, “Object tracking based on machine vision and improved svd algorithm”, International Journal on Smart Sensing and Intelligent Systems,vol. 8, No. 1, 2015, pp. 677-691.

[3] H.D.Yang, C.X.Wang, “Performance measurement of photoelectric detection and target tracking algorithm”, International Journal on Smart Sensing and Intelligent Systems,vol. 8, No. 3, 2015, pp. 1555-1575.

[4] Y.Q.Wang, C.X.Wang, “Computer vision-based color image segmentation with improved Kinect clustering”, International Journal on Smart Sensing and Intelligent Systems,vol. 8, No. 3, 2015, pp. 1707-1729.

[5] R.Singh, B.C.Pal, R.A.Jabr. “Statistical representation of distribution system loads using Gaussian mixture model”,IEEE Trans on Power Systems, vol. 25, No. 1, 2010, pp. 29-37.

[6] J.Lee, M.Park,“An adaptive background subtraction method based on kernel density estimation”,Sensors, 2000, pp. 12279-12300.

[7] K.Kim, T.H.Chalidabhongse, D.Harwood,“Real-time foreground background segmentation using codebook model”,Real-Time Imaging, vol. 11, No. 3, 2005, pp. 172-185.

[8] L.M.Hu,L.L.Duan, X.D.Zhang,“Moving object detection based on the fusion of color and depth information”,Journal of Electronics & Information Technology, vol. 36, No. 9, 2014, pp. 2047-2052.

[9] C.Stauffer, W.E.L.Grimson,“Adaptive background mixture modelsfor real-time tracking”,IEEE International Conference onComputer Vision and Pattern Recognition,Fort Collins, USA, June 1999, pp.246-252.

[10] A.Mittal, N.Paragios,"Motion-based background subtractionusing adaptive kernel density estimation", IEEE Conference inComputer Vision and Pattern Recognition,vol. 2, No. 2, 2004, pp.302-309.

[11] J.Leens, S.Piérard, O.Barnich, "Combining color, depth,and motion for video segmentation",LNCS, 2009, pp. 104-113.

[12] E.Mirante, M.Georgiev, A.Gotchey, "A fast image segmentation algorithm using color and depth map",IEEE 3DTV-Conference on the True Vision-Capture, Transmission and Display of 3D Video, Antalya, Turkey, 2011, pp.1-4.